

# Ensuring convergence in total-variation-based reconstruction for accurate microcalcification imaging in breast X-ray CT

Jakob H. Jørgensen, *Student Member, IEEE*, Emil Y. Sidky, *Member, IEEE*, and Xiaochuan Pan, *Fellow, IEEE*

**Abstract**—Breast X-ray CT imaging is being considered in screening as an extension to mammography. As a large fraction of the population will be exposed to radiation, low-dose imaging is essential. Iterative image reconstruction based on solving an optimization problem, such as Total-Variation minimization, shows potential for reconstruction from sparse-view data. For iterative methods it is important to ensure convergence to an accurate solution, since important image features, such as presence of microcalcifications indicating breast cancer, may not be visible in a non-converged reconstruction, and this can have clinical significance. To prevent excessively long computational times, which is a practical concern for the large image arrays in CT, it is desirable to keep the number of iterations low, while still ensuring a sufficiently accurate reconstruction for the specific imaging task. This motivates the study of accurate convergence criteria for iterative image reconstruction. In simulation studies with a realistic breast phantom with microcalcifications we compare different convergence criteria for reliable reconstruction. Our results show that it can be challenging to ensure a sufficiently accurate microcalcification reconstruction, when using standard convergence criteria. In particular, the gray level of the small microcalcifications may not have converged long after the background tissue is reconstructed uniformly. We propose the use of the individual objective function gradient components to better monitor possible regions of non-converged variables. For microcalcifications we find empirically a large correlation between nonzero gradient components and non-converged variables, which occur precisely within the microcalcifications. This supports our claim that gradient components can be used to ensure convergence to a sufficiently accurate reconstruction.

**Index Terms**—X-ray CT, breast CT, algorithm convergence, total variation, compressed sensing

## I. INTRODUCTION

**D**OSE reduction has gained considerable interest in diagnostic computed tomography (CT) in recent years [1]. The potential to employ CT for screening, where a large population fraction will be exposed to radiation dose and the majority of subjects will be asymptomatic, also motivates the interest in low intensity X-ray CT. Breast CT poses

a particularly challenging problem as the total exposure is restricted to the equivalence of two digital mammograms. Such a low X-ray dose can be achieved either by drastically reducing the intensity compared to a diagnostic-quality CT scan, or by reconstruction from sparse-view data.

Total-Variation (TV)-regularized image reconstruction exploits approximate sparsity of the spatial gradient of cross sections of the human body to compensate for reduction in data. TV-reconstructions have been shown to compare favorably with standard Filtered Back Projection from sparse-view data [2], [3]. We are investigating the optimal trade-off between low intensity views and sparse-view data for breast CT by means of TV-reconstruction [4].

The TV-reconstruction is obtained by solving a nonlinear optimization problem. A practical concern is that the extremely large systems in CT, where image arrays of  $10^9$  voxels are standard, are challenging to solve accurately in acceptable time. Complicating this issue is the fact that clinically relevant features are often very small—occupying only a few voxels. As result both global and pointwise convergence of an iterative reconstruction algorithm may have clinical impact. We demonstrate this issue in the present preliminary investigation, where we examine a realistic simulation of CT for breast cancer screening, and compare strategies for ensuring convergence to a sufficiently accurate TV-reconstruction.

## II. IMAGE RECONSTRUCTION BY TV-MINIMIZATION

We consider TV-regularized image reconstruction in order to exploit gradient sparsity to compensate for the few-view projection data. The present study works with the discrete-to-discrete imaging model,  $Au = b$ , see [5]. For reconstruction we consider the minimization problem

$$u_{TV} = \underset{u}{\operatorname{argmin}} f(u), \quad (1)$$

where

$$f(u) = \|Au - b\|_1 + \lambda \|u\|_{TV} \quad (2)$$

and

$$\|u\|_{TV} = \sum_j \|D_j u\|_2 \quad (3)$$

and  $D_j$  is a forward difference approximation to the image gradient at pixel  $j$ .

Instead of the more commonly used  $\ell_2$  norm for measuring data fidelity we use the  $\ell_1$  norm. TV-regularized  $\ell_2$  norm

This work is part of the project CSI: Computational Science in Imaging, supported by grant 274-07-0065 from the Danish Research Council for Technology and Production Sciences. This work was supported in part by NIH R01 grants CA120540, EB000225. The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

J. H. Jørgensen is with the Department of Informatics and Mathematical Modelling, Technical University of Denmark, Richard Petersens Plads, Building 321, 2800 Kongens Lyngby, Denmark (e-mail: jakj@imm.dtu.dk).

E. Y. Sidky and X. Pan are with the Department of Radiology, University of Chicago, 5841 S. Maryland Ave., Chicago IL, 60637, USA (e-mail: {sidky,xpan}@uchicago.edu).

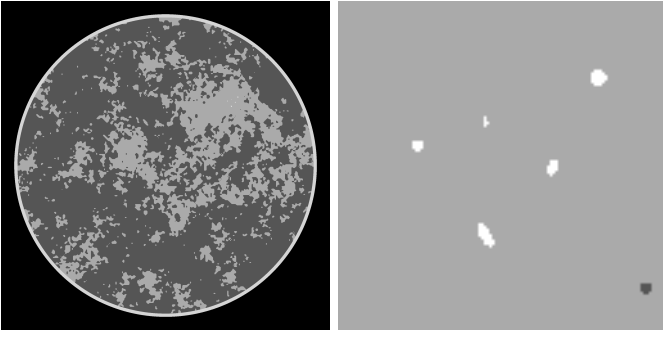


Fig. 1. Left: Original full breast phantom,  $2048^2$  pixels. Right:  $120^2$  pixel ROI around simulated microcalcifications. Gray level window:  $[0.9, 1.2]$ .

minimization is known to be contrast-reducing, in particular for objects of small scale [6], such as microcalcifications.  $\ell_1$  minimization does not remove this problem, but tends to reduce it [7].

Both terms in (2) are non-differentiable, and in order to apply standard gradient-based optimization algorithms we apply the standard smoothing trick of the replacements:

$$\sum_i \sqrt{\|D_j u\|_2^2 + \epsilon} \quad \text{replaces} \quad \|u\|_{\text{TV}}. \quad (4)$$

$$\sum_i \sqrt{|(Au)_i - b_i|^2 + \epsilon} \quad \text{replaces} \quad \|Au - b\|_1. \quad (5)$$

In our simulations we use  $\epsilon = 10^{-4}$ , which we found sufficiently small to prevent any change in visual appearance of the reconstructed image compared to using  $\epsilon = 0$ .

An important question is how well a TV reconstruction is capable of reproducing the salient image features, such as microcalcifications in the present case. Numerous studies demonstrate that of TV-reconstruction can produce clinically useful reconstructions, see e.g. [2], [3].

Our main question of interest in the present work arises when using an iterative algorithm to solve the TV minimization problem: When can we reliably stop iterating and accept the computed solution as a good approximation of the true minimizer to (2)? In other words, what is a good termination criterion?

Note that, in general,  $u_{\text{TV}}$  is biased compared to the original underlying image, and the size of this bias is parameter dependent, in particular the bias depends on  $\lambda$ . It is not our goal to select a well-suited  $\lambda$  here; we only consider the question of, given a choice of  $\lambda$ , how do the iterates approach the solution, i.e., the minimizer of (2)? We will consider two different choices for termination criterion for the iterative algorithm used for solving (1):

- 1)  $\|\nabla f(u)\|_2 < \tau$
- 2)  $1 + \cos \alpha < \tau$ ,

where  $\alpha$  is the angle between the gradients of each of the two terms in (2), see the original reference [3] for details, and  $\tau$  is a user-specified tolerance, where a smaller  $\tau$  leads to a more accurate solution. Both criteria correspond to theoretical optimality conditions [8] in the limit of  $\tau = 0$ .

For solving (1) we use a convergent, gradient-based optimization algorithm, which is optimal in a certain sense, see

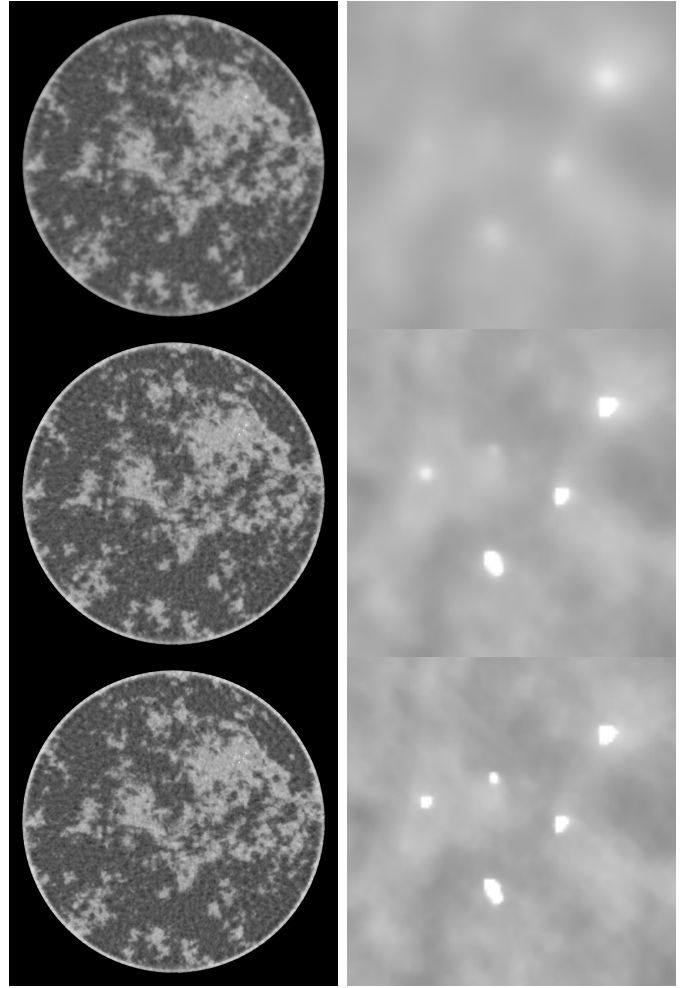


Fig. 2. Reconstructions of full image and ROIs. Top:  $\lambda = 2 \cdot 10^{-2}$ , middle:  $\lambda = 2 \cdot 10^{-3}$ , bottom:  $\lambda = 2 \cdot 10^{-4}$ . Gray level window:  $[0.9, 1.2]$ .

[9]. The algorithm was developed for TV-regularized  $\ell_2$  data fidelity, but is applicable to any smooth objective function, and we have found that it works well for solving (the smoothed version of) the problem in (1).

### III. BREAST CT MODEL

Breast CT imaging is being considered as a potential addition to mammography in screening for breast cancer. One particular indicator of breast cancer is formation of *microcalcifications*—very small, highly attenuating calcium deposits. For screening, low-dose imaging is pertinent to minimize accumulated X-ray dose, while accurate and reliable microcalcification shape and attenuation reconstruction may be important for detecting malignancy.

In the present work we use the breast phantom from [10] discretized on a  $2048^2$  pixel grid, as shown in Fig. 1 along with a  $120^2$  pixel region of interest (ROI) around a simulated cluster of microcalcifications, also discretized. Gray values in units of water attenuation are given in Table I. Note the fairly complex phantom structure, which makes the phantom semi-realistic, and at the same time poses a challenge for TV-based reconstruction, which tends to favor piecewise constant, “cartoon-like” images.

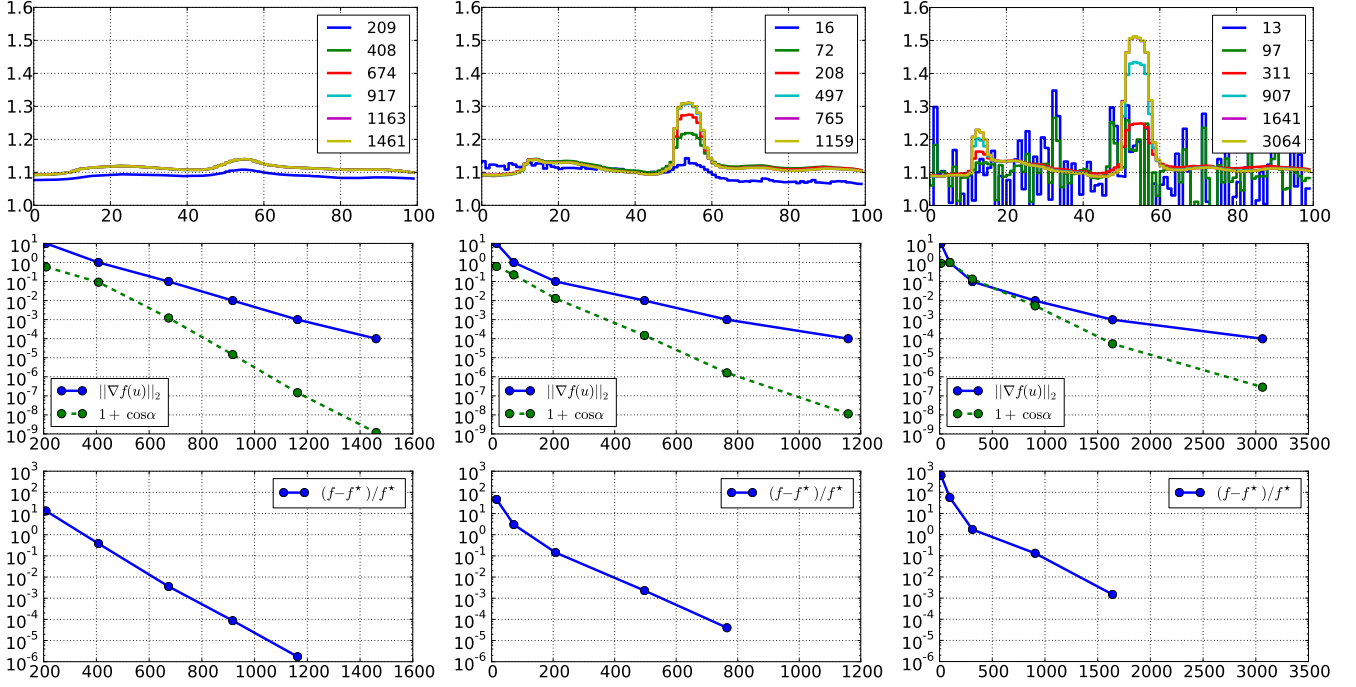


Fig. 3. Top row: Vertical profile through microcalcifications for iteration number corresponding to terminating iterations at  $\tau = 10^1, 10^0, \dots, 10^{-4}$  using criterion 1. Middle row: Values of the two convergence criteria vs. number of iterations. Bottom row: Convergence in objective function relative to the reference solution vs. number of iterations. Left column:  $\lambda = 2 \cdot 10^{-2}$ , center column:  $\lambda = 2 \cdot 10^{-3}$ , right column:  $\lambda = 2 \cdot 10^{-4}$ .

Tissue	Value
Fat	1.00
Fibro-glandular tissue	1.10
Skin	1.15
Microcalcifications	1.80 – 2.10

TABLE I

GRAY VALUES FOR BREAST PHANTOM, IN UNITS OF WATER ATTENUATION.

#### IV. NUMERICAL RESULTS

Different choices for the regularization parameter  $\lambda$  lead to very different solutions, and the question of how to choose a well-suited  $\lambda$  is important. However, our goal here is merely to demonstrate that very different convergence is observed for different choices of  $\lambda$ ; not to propose a certain  $\lambda$  over others. For that purpose we make three choices:  $\alpha = 2 \cdot 10^{-2}$ ,  $\alpha = 2 \cdot 10^{-3}$ , and  $\alpha = 2 \cdot 10^{-4}$ . We generate noise-free 64-view, 1024-detector-bin fan-beam data by forward projection (using a line intersection-based ray-driven projector) of the original discrete  $2048^2$  pixelized phantom with microcalcifications. We solve (1) with termination criterion 1 for  $\tau = 10^{-4}$  to obtain accurate solutions. The obtained reconstructions are shown in Fig. 2.

As expected, with increasing  $\lambda$  the reconstructed images becomes smoother, and the microcalcifications gradually become invisible. Only at  $\lambda = 2 \cdot 10^{-4}$  is the smallest microcalcification visible, so it is clear that we need to use a  $\lambda$  smaller than or equal to  $2 \cdot 10^{-4}$ .

We rerun the three reconstructions and store this time iterates along the way, at thresholds  $\tau = 10^1, 10^0, \dots, 10^{-4}$  for termination criterion 1. We use the most accurate iterate, at each  $\lambda$  as a *reference solution* for comparing the convergence

of the earlier iterates. We denote the reference solution by  $u^*$  and its value for the objective function by  $f^*$ .

In the top row in Fig. 3 we show reconstruction profiles through two of the microcalcifications (the two that are on the same vertical line) for each of the stored iterates, including the reference solutions. For the largest  $\lambda$  we see that the iterates converge to the reference solution very quickly: after 438 iterations i.e. at  $\tau = 10^0$  the solution is indistinguishable from the reference solution. For the middle  $\lambda$  we see a different behavior to which we will refer as *non-uniform convergence*: For the most part, the iterates converge to the reference solution rapidly, but precisely within the larger of the two microcalcifications, the iterates converge very slowly. For the smallest  $\lambda$  the non-uniform convergence is even more pronounced and only at the reconstruction stored before the reference solution we see no further improvement of the iterates. It seems natural that for even smaller values of  $\lambda$  we would see even more severe non-uniform convergence.

Our concern about non-uniform convergence arises from two facts: First, detecting non-uniform convergence can be very challenging as we will demonstrate. Second, if we are not aware of non-uniform convergence, we risk accepting a solution which is not yet converged everywhere. Such a reconstruction has much lower contrast than the true TV-solution, which will make it difficult to spot the microcalcifications. This can lead us to the, incorrect, conclusion that the TV-solution is not capable of reproducing microcalcifications faithfully, but in fact the lack of contrast in the reconstruction was a result of accepting a too early iterate returned by the iterative solver and not because of the TV-minimization problem itself.

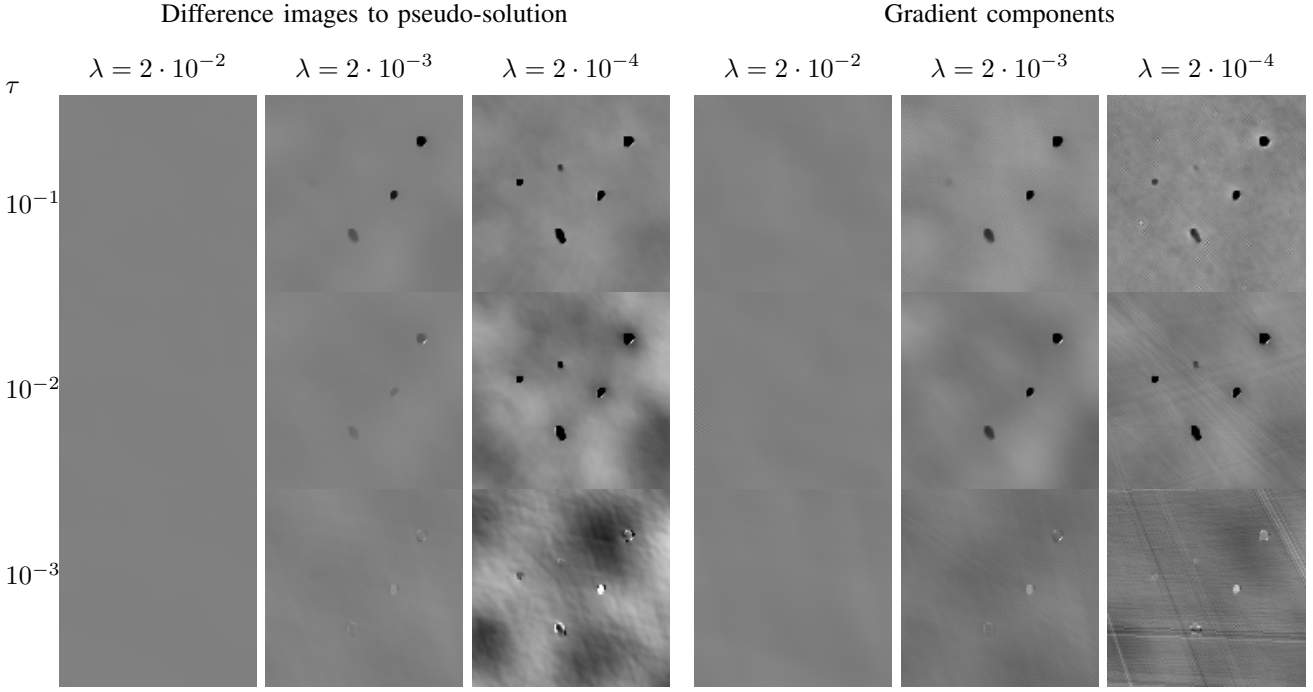


Fig. 4. Left: Difference images  $u - u^*$ . Gray level windows: Top:  $[-1 \cdot 10^{-1}, 1 \cdot 10^{-1}]$ . Middle:  $[-2 \cdot 10^{-2}, 2 \cdot 10^{-2}]$ .  $[-2 \cdot 10^{-3}, 2 \cdot 10^{-3}]$ . Right: Gradient components  $\nabla f(u)$ . Gray level windows: Top:  $[-1 \cdot 10^{-4}, 1 \cdot 10^{-4}]$ . Middle:  $[-1 \cdot 10^{-5}, 1 \cdot 10^{-5}]$ .  $[-1 \cdot 10^{-6}, 1 \cdot 10^{-6}]$ .

To illustrate that detecting non-uniform convergence is challenging, we compare the use of termination criteria 1 and 2. In the middle row of plots in Fig. 3 we have plotted the values of the two criteria vs. the number of iterations used for the considered iterates. Furthermore, in the bottom row of plots we show the convergence in terms of objective function value relative to the reference solution vs. the number of iterations. The difference in function value relative to the reference solution acts as an indicator of the accuracy of the iterate. Now imagine that we use  $\tau = 10^{-3}$  as our convergence criterion. With criterion 1 and the largest  $\lambda$  we find an  $(f - f^*)/f^*$  of approx.  $10^{-6}$ , which we consider to be very accurate. However, for the two smaller values  $\lambda$  we find  $(f - f^*)/f^*$  of approx.  $10^{-4}$  and  $10^{-3}$ , indicating much less accurate reconstruction. A similar trend can be observed for termination criterion 2. The different accuracies obtained confirm our observations from the profile plots in the top row. We note that we are able to detect non-uniform by comparing the final function value differences in the plots in the bottom row. However, in practice, we do not have access to the true solution or a reference solution, as we would like keep the number of iterations low. We do have access to the values of the termination criterion functions, but as can be seen by inspecting the middle row of plots, we cannot trust that a using a fixed  $\tau$  will provide a uniformly converged reconstruction.

## V. GRADIENT COMPONENTS

As a first step towards a more reliable convergence criterion we wish to point out a connection that can possibly exploited. The two considered convergence criteria both involve the gradient of the objective function  $f$ . However, as we saw, they do not clearly show that a few pixels have not yet

reached convergence. We believe this is due to computing a single number from the full gradient for comparing with a  $\tau$ , thereby “averaging out” the differences between the individual components of the gradient. Many small gradient components will tend to hide the presence of a few larger ones. We propose instead to monitor the full objective function gradient  $\nabla f(u)$  during the iterations.

In the right half of Fig. 4 we display as an image the ROI gradient components of the objective function, for the iterates obtained with  $\tau = 10^{-1}, 10^{-2}, 10^{-3}$  for each of the three choices of  $\lambda$ . In the left half of Fig. 4 we show the corresponding ROI difference images  $u - u^*$  between the iterates and the reference solution image. Note that with decreasing  $\tau$  we are making the gray level windows narrower to emphasize small components.

For the largest  $\lambda$  both difference images and gradient components appear to be zero (in the chosen gray level window) for all three choices of  $\tau$ . This agrees well with our observation that the iterates converged rapidly to the reference solution, so we should precisely expect very small gradient components everywhere here.

For the smaller choices of  $\lambda$  we observe a highly non-uniform nonzero gradient component pattern for both the difference images and the gradient components, with large (negative) components exactly at the microcalcifications. The gradient components are negative, which agrees with the variables still growing as seen in the profiles in the top row of Fig. 3. For the more accurate reconstructions the microcalcification pixel values in the difference image and gradient components remain distinct while their magnitude approach zero.

There is a clear correlation between the difference images and the gradient components, indicating a close connection. This suggests the possibility for ensuring local convergence in

the microcalcifications by means of monitoring the gradient components.

At the most accurate solution, while the microcalcification pixels are still visible in difference images and gradient components, the intensity is of the level of the background. This leads us to the conclusion that at this point the iterate has converged, and we can reliably accept it as an accurate solution.

## VI. DISCUSSION

Note that we are only able to monitor the convergence using the difference images, since we computed the reference solution. In practice, we wish to monitor convergence at any given iteration without a much more accurate reference solution. The gradient components are readily available during the iterations, and as our simulation shows, they can be used to monitor non-converged pixels.

We are investigating strategies other than visual inspection of the gradient components for a quantitative convergence criterion. For instance by forcing  $\max_j |(\nabla f(u))_j|$  below a appropriately chosen threshold  $\epsilon$ , all gradient components will be smaller than  $\epsilon$ , thereby ensuring global convergence. When applying a single number based convergence criterion such criteria 1 and 2, the fact that the majority of the variables are at optimum can conceal by averaging out the contributions from the few variables that are not. The rationale in forcing all gradient components below  $\epsilon$  is that small areas of non-convergent variables will prevent termination of the algorithm. A different approach would be to exploit the spatial structure in the nonzero gradient components, e.g. by not terminating iterations until no spatial correlation is present.

## VII. CONCLUSION

We have conducted a preliminary comparative investigation of convergence criteria for ensuring accurate reconstruction of microcalcifications in breast CT. We have demonstrated that the nonzero gradient components can be used to monitor the regions of non-converged variables and thereby preventing termination of the optimization algorithm before global convergence is reached.

Accepting a reconstruction which is not globally converged may have clinical significance, for instance, as in the example given, by providing insufficient contrast for detecting the microcalcifications.

The use of the objective gradient in a convergence criterion is well-known, at least the use of the norm of the gradient. Explicit use of the individual gradient components for monitoring local convergence for small objects such as microcalcifications has not, to the best of our knowledge, been studied before. An interesting direction for future work is to apply the approach to other optimization based reconstruction techniques.

## REFERENCES

- [1] C. H. McCollough, A. N. Primak, N. Braun, J. Kofler, L. Yu, and J. Christner, "Strategies for reducing radiation dose in ct," *RADIOLOGIC CLINICS OF NORTH AMERICA*, vol. 47, pp. 27–40, 2009.
- [2] E. Y. Sidky, C.-M. Kao, and X. Pan, "Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT," *JOURNAL OF X-RAY SCIENCE AND TECHNOLOGY*, vol. 14, no. 2, pp. 119–139, 2006.
- [3] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *PHYSICS IN MEDICINE AND BIOLOGY*, vol. 53, no. 17, pp. 4777–4807, 2008.
- [4] J. H. Jørgensen, P. C. Hansen, E. Y. Sidky, I. S. Reiser, and X. Pan, "Toward optimal X-ray flux utilization in breast CT," in *Proceedings of the 11th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, Potsdam, Germany, 2011, available from <http://arxiv.org/abs/1104.1588>.
- [5] H. H. Barrett and K. J. Myers, *Foundations of Image Science*. Hoboken, NJ: John Wiley & Sons, 2004.
- [6] D. Strong and T. Chan, "Edge-preserving and scale-dependent properties of total variation regularization," *INVERSE PROBLEMS*, vol. 19, no. 6, pp. S165–S187, 2003.
- [7] T. F. Chan and S. Esedoglu, "Aspects of total variation regularized L-1 function approximation," *SIAM JOURNAL ON APPLIED MATHEMATICS*, vol. 65, no. 5, pp. 1817–1837, 2005.
- [8] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. New York: Springer, 2006.
- [9] T. L. Jensen, J. H. Jørgensen, P. C. Hansen, and S. H. Jensen, "Implementation of an optimal first-order method for strongly convex total variation regularization," *BIT Numerical Mathematics*, to appear, 2011, preprint available from: <http://arxiv.org/abs/1105.3723>.
- [10] I. Reiser and R. M. Nishikawa, "Task-based assessment of breast tomosynthesis: Effect of acquisition parameters and quantum noise," *Med. Phys.*, vol. 37, pp. 1591–1600, 2010.